

出國報告（出國類別：國際會議）

參加 2014 年 SAS 全球論壇會議
(SAS GLOBAL FORUM)
出國報告

服務機關：行政院主計總處

姓名職稱：任晶蓉分析師

派赴國家：美國

出國期間：103/3/21~103/3/27

報告日期：103/05/27

摘要

SAS 全球論壇會議 (SAS GLOBAL FORUM) 係由非營利組織 SAS 軟體全球用戶社群 (SAS GLOBAL USERS GROUP) 舉辦之年度國際研習會，主要探討 SAS 在金融、醫藥衛生、生產、運輸、通訊、科學研究、政府及教育等各領域之應用實務與最新研究，讓全球 SAS 使用者藉此會議相互交流、獲取 SAS 新知及解決方案、強化 SAS 使用能力與技巧、激盪創意。

2014 年 SAS GLOBAL FORUM 會議期間邀請各國學者專家、產業實務專家、研究員、顧問、資深使用者分別就 **1.專業技能 (Skill Level)** **2.工作角色 (Job Roles)**：分析師、商業分析師、IT 人員、統計人員等 **3.行業 (Industry)**：製造業、銀行業、通訊業、教育界、政府部門、醫療照護、保險業等 **4.SAS 產品 (Product)**：網格運算管理、商業智慧分析、詐欺偵測管理、企業風險管理、資料探勘、視覺化分析、預測等，四個面向以 **Breakout Session**、**Hands-On Workshop**、**Quick Tips**、**Panel**、**Live Stream** 方式開辦 **384 場議程**。

巨量資料引發資料探索與關鍵變數挖掘，對總處埋首於多年公務統計或調查統計資料之統計分析人員們，感受其工作的辛勞與重要性。如何在此排山倒海而來的資料中找到重要問題答案、協助決策，需仰賴資料科學家的協助，因此，強化總處既有分析人員的功力，由分析家躍進為資料探勘家是可以考慮的；若能再導入高效分析軟體或工具，對提昇總處的競爭力必有幫助。

目 次

壹、目的.....	1
貳、過程.....	1
參、會議內容摘錄.....	6
一、巨量資料分析義涵與發展概況	6
巨量資料之分析.....	7
巨量資料之下新興的 IT 技術.....	7
Hadoop—巨量資料分析平台的最佳選擇之一.....	8
二、統計方法與資料分析.....	9
資料探勘 (data mining)	10
三、SAS 軟體在巨量資料分析的做法.....	12
SAS Analytics Pro (本處有其產品)	12
SAS Enterprise Guide (本處有其產品)	13
SAS 巨量資料視覺化分析平台.....	15
SAS 高效能運算分析.....	16
SAS Mobile BI.....	17
肆、心得與建議.....	19
一、資料科學與資料科學家.....	19
二、大資料需要大分析 (Big Data need Big Analytics)	21
三、將資料當做策略性、核心資產來管理	23
四、地理視覺化報表.....	23
伍、結語.....	25
陸、參考資料.....	26
柒、附錄.....	27

壹、目的

鑑於近來政府資料開放（Open Data）及巨量資料（Big Data）分析漸為我國政府重要資訊發展範疇，為提升巨量資料應用與強化統計分析服務效能，參加非營利組織 SAS 軟體全球用戶社群（SAS GLOBAL USERS GROUP）於美國華盛頓特區舉辦之 2014 年 SAS 全球論壇會議（SAS GLOBAL FROUM），會期自 3 月 23 日至 26 日為期 4 天[1]。

希望藉由本次會議研習、經驗交流，有助強化本總處統計分析、資料探勘與資料應用等技術能量。

貳、過程

此次參加的 SAS 全球論壇會議 — **Potential of One, Power of All**，是由 SAS 全球用戶社群每年於美國不同地點舉辦之會議，今年於 3 月 23 日至 26 日在華盛頓特區南部 8 英里（13 公里）的國家港灣水濱蓋洛度假村酒店及會議中心（Gaylord National Resort & Convention Center）舉行，參與人次超過 4500 位，打破歷年記錄，不但是 SAS 軟體公司的年度盛典，也是全球 SAS 使用者時時關切的焦點，以下是會議的官網：
<https://support.sas.com/events/sasglobalforum/2014/>

開幕式

🌈 SAS 總裁 Jim Goodnight 博士提及 2014 年大眾頻頻接觸到雲端 (Cloud)、社交媒體 (Social Media)、社交社群 (Social Network)、行動化 (Mobility)、大資料 (Big Data)、Cyber Fraud 等名詞，與這些辭彙關聯的分析 (Analytics)、決策 (Make Decision) 已經遍佈全世界，使得人們不得不體認資料探索與關鍵變數發掘，在資料解析過程中扮演不可或缺的角色。SAS 不斷地與各行各業共同發展商業資料分析與預測技術，應用領域涵蓋政府的經濟決策與企業的決策支援應用等，因應現代化 (Modernization) 的 Grid Manager、Visual Analytics、Cloud Analytics、Text Analytics 等技術就是為解決大資料分析所在意的高效能與視覺化而興起的解決手法。SAS Lasr 分析伺服器打造優越的運算環境，不但加速分析運算速度並可與 Hadoop 分布式檔案系統完美結合，讓企業能兼顧龐大資料儲存與平行運算。在因大數據成長與應用層面所迫切需要的資料科學家的養成上也不缺席：除了引入 SAS University Edition Software 外也積極參與各行各業的資料解析方法建置。

🌈 **SAS Fraud & Security Solution Overview**—Stu Bradley (Director of Financial Crimes Practice with SAS)

◆ SAS 面對近年來盛行之 Cybersecurity and Fraud 事件以 Visual Analytics、Fraud Detection and Prevention、Consistent

Monitoring 整合詐騙偵測、警示管理、社交網路分析及案件管理的技術基礎結構，即時迅速的地圖上，以可視化的方式及早偵測可疑活動。得以掌控與偵測位於任何接觸點的詐騙。

 **Modernizing Analytics**—Jim Davis(SAS Senior Vice President and Chief Marketing Officer)

- ◆ 現代化 (modernize) 是因應現代行為或需求所建置的設備、想法或解決手法。
- ◆ SAS 在因應巨量資料 (Big Data)、分析 (Analytics)、快速決策 (fast-make-decision) 等現代化需求所作的解決方案包括商業智慧 (BI) 平台、Enterprise Guide、Hadoop、可視化 (Visual)、導入記憶體 (In-Memory)、雲端 (Cloud)、LASR Server (高速大量資料分析伺服器無不竭盡所能滿足客戶期望)。

Keynote Speech :

🚩 「**Remembering Not to Forget**」—Nelson Dellis (USA Memory Champion and Founder and CEO of Climb for Memory) ; 3/23

◆ 如何經由長期訓練成為記憶專家

🚩 「**Big Data, IT Issues**」—Thornton May (Futurist and Executive Director at the IT Leadership Academy)

◆ 以集體討論的方式結論出

1. The hero of the next age will be the person who effectively "Monetizes Meaning" Via Mastery of Big data/Big Analytics

2. The "with-data" aggressive & confident problem solver will be to the Age of Big Data what the Astronaut was to the Space Age.

🚩 「**Statistically valid suggestions for 2015**」—Sebastian Wernicke (Statistics Whiz)

◆ Schedule problem are easy.

◆ Providing 50% in-database performance

◆ Highlight: language quality

TECH Talk :

-  **SAS Visual Statistics and SAS In-Memory ,**
-  **Statistics for Hadoop**
-  **SAS and the Cloud,**
-  **SAS Visual Analytics Graph Builder, SAS/IML® and R**
-  **SAS Studio,**
-  **SAS Add-In for Microsoft Office,**
-  **Using SAS Text Analytics to Analyze Medical Reports,**
-  **New ODS destinations: HTML5, EPUB, PPT, EXCEL**

議程：

共計 384 場[2]。

參、會議內容摘錄

一、巨量資料分析義涵與發展概況

當資訊計量單位由 MB (MegaByte=1000Byte)、GB (GigaByte)、TB (TeraByte)、到 PB (PetaByte,)、EB (ExaByte)、ReheB、YottaB [3]時，意味者資訊爆炸時代的來臨。大資料 (Big data) [4]，或稱巨量資料、海量資料、大數據，是指所涉及的資料量規模巨大到無法透過人工，在合理時間內達到擷取、管理、處理、並整理成為人類所能解讀的資訊。數據資料量與種類快速成長的驅動力，主要是來自於越來越多且便宜的電腦儲存設備，及資訊持續從各種來源被廣泛收集(感應器、測量器、追蹤器等設備以每年 30%以上的成長速度激增)，且 90%以上的資料將會以數位的方式來處理[5]；加上社群媒體、行動網路及物聯網的推波助瀾。數據資料的變化，包括多媒體資料檔案高量成長、個人數位裝置的暴增，導致使用與互動方式的改變：大型綜合巡天望遠鏡 (Large Synoptic Survey Telescope, LSST) 計畫，每晚可收集約 20 TB (1 TB=1000 GB) 的天文資料；網路公司，像 Google 每天得處理超過 24 PB (1 PB=1000 TB) 的資料；Facebook 每小時會收到超過 1 千萬張新照片、30 億次的留言；YouTube 用戶每秒上傳的影片總長度超過 1 小時；隨者資料愈趨龐大，花費在彙整與分析的時間愈多，導致愈無法再度重覆執行或溯及既往的追溯，以及資料流動過快，不知如何處理的非結構化資料等議題出現的窘境。

資料獲取及創建的技術來源既多且廣，且資料之間的關係繁複，導致創建資料的速度、分享也變得更加容易，廉價的儲存成本，更讓資料保存的時間也相對拉長，成為驅動巨量資料成長的重要因素。根據國際數據中心 IDC 統計，數位世界的資訊容量不斷成長，將會從 2009 年的 0.8ZB，在 2020 年成長到 40ZB (相當於 1 兆 GB)，等於每 15 秒就成長 1PB，年複合成長率高達 40%，在爆炸性的資料增長過程中，結構性資料的成長相當緩慢，非結構性的資料，包括視訊、網頁、

智慧型手機、消費資料、位置資料、財務服務資料，以及社會媒體資料等，至少有 80%的數據，屬於非結構化資料。

爆炸性的資料流動令使用者頻頻更新資料，且資訊系統也日趨複雜，促使巨量資料 (Big Data) 的分析應用成了當前最重要的課題。對於資料量的爆增，企業除了需要從龐大且複雜的資料中整理出有效的訊息之外，大資料正驅動我們到一個轉捩點：及時的資料形成、多樣的資料型態、資料的量變大、使用者的期望、硬體與儲存的成本經濟、行動性。

巨 量 資 料 之 分 析

巨量資料的分析多用在三種行為，分別是評估、即時決策以及預測。企業資料就像冰山一樣，看得到的部分往往只有浮在表面的 10%，看不到的約 90%，過去習慣處理的資料，其實只佔海量資料的 15%，往往是表面資料出問題時，才會去檢視後台的底層資料 (Row DATA)。這是一種被動解決問題的態度，企業如果能在平常就將底層資料分析好，才有可能事先找到機會，而非事後解決問題。看似垃圾的龐雜資料，只要經過巨量資料技術分析，透過細微的觀察、分析及萃取，從資料中發掘有價值的知識，才算是發掘到金礦。想從巨量資料煉金，第一種思維是用方法去找問題，第二種思維是依問題來找方法，前者是提供眾多產品，後者則是專精特定領域，兩者各有價值。

巨 量 資 料 下 新 興 的 I T 技 術

在資料增長過程中，結構性資料的成長往往相當緩慢，但非結構性的資料，包括視訊、網頁、智慧型手機、消費資料、位置資料、財務服務資料，及社群媒體資料等，至少有 80%的數據，屬於非結構化資料，已經超越傳統資料庫或現有資料管理工具能夠處理的範圍。目前的資料庫解決方案，主要是用來設計儲存結構化資料，能對已知問題的回答速度進行優化，並且資料架構本身往往就決定了內容形式，對於新資料型態與新問題，都有適應上的困難，加上擴展成本高昂，

企業勢必得尋求不同以往的資料處理解決方案，才能面對爆炸性的資料增長。

如何優化資料，方便且容易的搜尋到所需要的資訊，也變得更加困難，因此各種新技術也如雨後春筍般冒出來：IN-Memory Computing、NOSQL DB、Visual Discovery、Grid and Clustering、Hadoop、MapReduce、EDW、Scale-out、Streaming Data Technology、Predictive Analytics 等等。

H a d o o p – 巨量資料分析平台的最佳選擇之一

大資料必須藉由電腦進行統計、比對、解析方能導出客觀結果，但因資料巨大幾乎無法使用大多數的資料庫管理系統處理，而必須使用「在數十、數百甚至數千台伺服器上同時平行運行的軟體」。最早碰到巨量資料問題的業者就是 Google，根據 Google 在 2006 年發表的資料所建構而成的 Hadoop 資料儲存與計算平台，就成為有意強化巨量分析的企業 IT 架構，最適合的選擇之一。

Hadoop 是一種分散式的資料處理架構，使用者可以在不了解分散式底層細節的情況下，開發分散式程式。Hadoop 尤其擅長複雜的分析，可充分利用集群的威力，高速運算和儲存，並有著高容錯的特點。Hadoop 更實現了分散式檔案系統（Hadoop Distributed File System；HDFS）架構，可以部署在低廉的硬體上，並且可提供高傳輸率，來查詢應用程式的資料，適合有著巨量資料集的應用程式使用。

由於 Hadoop 主要是適用於非結構化或半結構化資料，與適用於結構化資料、歷史性交易資料的關聯式資料庫，其實有相輔相成的效果，Hadoop 與資料倉儲的應用特性並不同，所擅長處理的資料類型也不同，在 Hadoop 的實務應用面，絕對不可能替代既有的資料倉儲。想要駕馭巨量資料，企業必須要能針對特別龐大的資料進行分析，擁有能處理密集數據的工作負載能力，並保證無限量的資料儲存能力，因此，巨量資料解決方案的需求，必須要具備成本優勢來管理資料的數量、創建速度與種類，並能夠處理與分析極大、複雜的資料，同時快速地進行，

且能彈性的適應環境變化與新資料的類型，企業 IT 架構勢必也要因應巨量資料，而有所變革。

二、統計方法與資料分析

現代的企業經常需要蒐集「大量資料」或「多維度資料」，包括市場、客戶、供應商、競爭對手、及未來趨勢等重要資訊。但資料超載與無結構化，使得企業決策單位無法有效利用現存的資訊，若能透過有效的資料探索技術或預測型資料探索，從大量的資料中採集不同的資訊與知識出來，作為決策支援之用，必能產生企業的競爭優勢。

大資料應用興起，除了傳統的報表、OLAP 分析，也開始有企業運用數學、統計、機器學習等方式來分析企業既有資料與外部資料，透過數據來發掘企業還沒發現到的問題，預測未來趨勢將不再是夢想。

統計是從複雜數據中萃取出有用資訊的學問，因此在分析巨量資料這件工作上，便扮演了舉足輕重的角色。統計技術與非統計技術的差異，主要在於是否考量隨機因素。母數如果太大，就需要統計技術，透過機率分配，用樣本回推母體。一般人所熟知的統計，多半是敘述統計，非統計專業的人也可以去做，特色是可以使用全部資料，但是要花費大量的金錢、時間及人力。

因此，可以從較小的樣本反推母體特徵的推論統計，也開始得到重視，所涉及的風險有抽樣誤差、決策錯誤，但在統計理論中，均能將這些風險控制降至最小甚至可忽略。雖然從學理來說，統計抽樣只要有一定的樣本，而且樣本身沒有問題就具有代表性，但相較之下，大多數人寧可相信從完整的資料所分析的結果比抽樣的效果更精確。面對巨量資料，有無必要使用全部資料，是許多欲在資料中挖掘有用資訊時可以思考的議題。會有這種想法起因於成本，而且除了有形的金錢成本外，時間及人力資源等無形成本，也都需要一併考量；其次是精確度會受到多少影響，造成的決策風險會有多高，最後則是軟硬體方面的限制，是否真的有足夠的能力使

用全部資料，進行統計分析。常見的資料分析類型為例，使用頻率最高的就是計數，也就是個數計算，如某篇網路文章有幾個人按讚，有 70% 以上的分析，都屬於這一類分析。其次是敘述統計及一般統計分析，如平均數、標準差、比例、迴歸、實踐設計、時間數列、多變量分析等，大約佔 20%。

資料探勘 (data mining)

資料探勘是在探討解析大資料的方法，是一個從複雜資料中挖掘，被隱藏且具有價值之型式與關係 (hidden pattern and relationship) 的過程。資料採礦的功能：

- (1) 分類 (Classification)
- (2) 預測 (Prediction)
- (3) 關聯分組 (Affinity Group)
- (4) 群集化 (Clustering)
- (5) 敘述 (Description)

其實是一種進階統計技術，可以用自動或半自動方式，將人力無法或忘了注意的資料找出來。資料採礦領域最常用的四大技術有：

- (1) 分類 (Classification)：預測資料所屬的類別
- (2) 預測 (Prediction)：預測資料所對應的數值
- (3) 分群 (Clustering)：發掘資料間的內部結構
- (4) 關聯 (Association)：找出那些事件常常一起出現

應用在實務方面常見如下：

(1) 描述型資料探勘，目標在找出趨勢，所用的分析技術：

(A). 推估 (**Estimation**)：預測資料所屬的類別，根據既有連續性數值之相關屬性資料，以獲致某一屬性未知之值。

(B). 關聯分組 (**Affinity Grouping**)：找出那些事件常常一起出現，從所有物件決定那些相關物件應該放在一起。

(C). 同質分組 (**Clustering**)：發掘資料間的內部結構，將異質母體中區隔為較具同質性之群組 (**Clusters**) 關聯分組。

(2) 預測型資料探索技術 (分類及預測)：預測資料所對應的數值，目標在預測/估計及異狀偵測。

資料探勘的應用俯拾即是，如應用在消費市場就有購物籃分析，用來作為行銷活動規劃；客戶關係管理則可以將資料探勘用來尋找潛在客戶，提升客戶終生身價，進而保持客戶忠誠度；其他如保險詐欺偵測、客戶信用風險評比、電信盜打，甚至 **NBA** 球賽球員攻守強弱分析，也都用得上資料探勘。

一般資料採礦的建模方式大都遵循 **CRISP-DM** 的標準流程:由商業理解、資料理解、資料準備、建立模型、效能評估、上線實施。為了達成 **CRISP-DM** 的各個階段，許多廠商紛紛推出許多資料採礦的工具，較為知名的有 **IBM Intelligent Miner**、**SAS Enterprise Miner**、**SPSS Modeler** (已被 **IBM** 所併購)。

在 Data Mining 的整個運作過程，資料前處理是最重要的一個步驟，資料前處理得當，對後續資料探勘過程是否有效和質量的好壞有重要的影響，其重要性和所花費的時間，均位居其他步驟的首位。SAS 工具在這個領域確為翹楚，由外部資料的資料編碼、資料轉換、資料精簡、記錄精簡、域值精簡、欄位精簡到資料選擇、數據整合、數據過濾、資料清理、錯誤值、離群值、空值、欄位擴充等等無不提供各種方法以達目的地。

資料探勘技術雖然早已成熟，採礦軟體多能提供建立探勘模型所需的元件，然而資料分析人員需將資料採礦流程及分類的詳細步驟熟記於腦海中，再根據腦海中的記憶一步步完成分類模型的建置，故優秀的資料分析人才是資料採礦成敗的關鍵。

三、SAS 軟體在巨量資料分析的做法

◆ SAS Analytics Pro（本處有其產品）

在這個資料即一切的時代，企業、政府或學術界，組織無時無刻被資料壓得喘不過氣，挑戰在於能否達到「隔天就分析完畢並以可視化的方式呈現結果」。資料來源極為龐雜，多數的統計軟體無法處理：此外，為配合政府法規與標準，必需採用越來越先進的統計方法，進行可驗證和記錄的分析。SAS Analytics Pro 結合 SAS 最熱門的三種軟體產品：Base SAS、SAS/STAT 與 SAS/GRAPH 為一完備的整合式套裝工具集，讓使用者對資料進行處理、儲存、分析，將資料轉換為具意義的實用資訊，以說服力強的視覺化效果呈現，協助決策者快速瞭解問題，並可在桌上型電腦上布署與執行，不論是分析家、研究人員、統計人員、工程師或科學家，可以用簡單明瞭的方式瀏覽、檢驗與呈現資料，更可以用各種格式發布自己的成果。

SAS Analytics Pro 提供具高度彈性且便於擴充的第四代程式設計語言，語法簡單易學，包含可協助進行程式設計的數百種語言元素與功能，涵蓋由資料擷取、格式化與清除，到資料分析、報告及資訊除存，並預先撰寫的 SAS 程序以處理許多常見的作業，包括資料處理與管理、資訊儲存、統計分析與報表撰寫，這些程序封裝並提供重要的功能，只需幾個簡單的指令即能執行，讓程式社計人員更具效率及生產力。

- **SAS/BASE**：一個專門設計之可擴展、整合的軟體環境，滿足讀取資料、轉換資料和編製告等工作，支援所有主要的計算平台，可顯著減少程式編撰與維護時間，讓使用者能快速的產製決策者偏愛格式的結果與報告。
- **SAS/STAT**：提供專業化且廣泛的統計分析方法及程序，從傳統的描述型分析與預測、建模。到精密方法與可視化技術，是企業級的分析需求工具。可處理大部分統計分析，計有變異數分析、迴歸分析、類別資料分析、多變量分析、存活分析、心理計量分析、長期資料分析、集群分析、無母數分析、貝氏分析，並有正規化、線型化與非線型化等變數轉換功能。這些功能均可產生可重覆程式碼，便於文件建立與驗證，從而符合公司與政府規範；並且 SAS 公司在開發分析軟體方面有 30 年以上的經驗，在提供高級可靠結果方面有很高的學術聲望。
- **SAS/GRAPH**：提供可視化及圖形呈現物件，可繪製豐富的圖表及圖形，建立客製不同的圖形類型，並可與其他資料聯結製作靜態或交互式可視內容；更透過 SAS ODS 輸出交付系統，將圖形結果以 ActiveX control、Java Applet、GIF、PNG、JPG 多樣面貌呈現在網頁瀏覽器中或發送至報表、或電子郵件。

◆ SAS Enterprise Guide（本處有其產品）

- 整合有效率且友善的圖形化介面，以點選式選單與精靈驅動的工具，提供程式設計人員、統計與分析人員等符合其工作需求

的自助式環境，以引導的方式，執行簡單或複雜的統計、分析或報表編製，一目了然的 SAS 資料與外部資料存取，可將最新的結果匯至其他 Windows 伺服器應用程式或使用者，讓管理部門得以制定正確的決策。以集中化及以角色為基礎的安全性管理機制，管理組織內資料的存取確保使用者擁有合適的權利，尚有程序流程圖工具，讓使用者清楚專案內容與步驟，支援許多立即可用的分析任務（task）：如預測、相關性分析或精靈（計 90 餘個），讓使用者輕鬆使用特定功能即可自動產生程式碼，初學者或進階使用者可省下程式設計時間。更甚者，客製化精靈製作可讓進階使用者結合精靈與程式設計，製作特定且自訂的任務精靈，引導較資淺的使用者一步步完成較複雜的程序，延伸產品的核心功能並可與網路上其他 Enterprise Guid 使用者分享，或者可透過其他的 SAS 商業智慧應用程式與 Microsoft Office 分享。

- **圖形化使用者介面**—直覺且具有彈性的精靈可提供報表製作、分析與資料存取，伴隨產生記錄檔，警訊及錯誤處理，結果可以 HTML、RTF、PDF、SAS 報表及文字檔輸出，可將大部分的結果另存為 SAS 資料集，與其他任務（task）一併用於後續分析，也可製作為 ActiveX 動態圖像之 Jav Applet、GIF 或 JPEG，並有直覺式程序流程圖介面，可一目了然各種工作程序。
- **資料存取與管理方面**，不但擅長存取任何 SAS 支援的資料類型，並可透過 ODBC、OLE DB、OLAP 等功能存取原生的 Windows 資料類型（計有 Microsoft Word 文件、Excel 檔、HML、Access、Paradox、TXT 檔、ASC 檔、TAB 分隔檔及以逗號分隔之 CSV 檔），OLAP 分析器提供專門且可進階分析儲存於 OLAP 資料方塊（cube）的資訊，並能瀏覽擷取多維度資料，透過進階統計程序潛進行資料探勘分析。
- **圖形化查詢產生器（Query Builder）**，可讓使用者在沒有 IT 人

員的協助下，建立更新各種查詢或子查詢（可預覽產生的查詢程式碼，藉以學習或強化查詢語法的能力），連結資料庫表格，即使非 SQL 專家也能輕鬆操作資料，完成資料轉置、排序、標準化等資料整理工作，並可存取其他 SAS 產品所建立的資訊。

- **報表製作與分析任務**的廣泛與周延，舉凡清單式報表、摘要式統計表、單因子次數統計表（one-way frequency table）、關聯表、圖形方面則有區域圖、長條圖、盒型圖、氣泡圖、環形圖、折線圖、地圖圖表、圓形圖、雷達圖散佈圖、曲面圖及等高線圖等，分析任務有變異數分析 ANOVA 預測模型（t 檢定、單因子 ANOVA、無母數單因子 ANOVA、線性模型及混合模型）；迴歸模型（線性迴歸、羅吉斯迴歸、非線性及廣義線性模型）；多變量關係模型（群集分析、因子分析、主成分分析、正準相關及判別分析）；存活分析；功能分析；管制圖；時間序列；表格分析等不勝枚舉。

◆ SAS 巨量資料視覺化分析平台

視覺化及高效能分析巨量資料的手法，搭配 SAS 合作夥伴（EMC Greenplum 或 Teradata）的 Database appliance 解決方案，利用導入記憶體式資料庫（In-Memory Database）的運算技術，迅速又準確的同時支援多位使用者(不論是決策者、分析師、統計人員或資料科學家)進行資料探索或分析，再經由網路或行動裝置以便即時發佈結果。IT 人員在完成平台建置後可經由資料架構師以一致性規範為不同使用者準備資料，留給使用者自行運用的空間且又可掌控資料公開程度、下載流量。

- **資料視覺化介面**：不論資料多寡或繁簡，Web 型態的互動式資料探索模式，適用各種使用者在不需仰賴 IT 部門下取用資料。自動化圖表會自動依據使用者欲分析的資料項目選取最適合的圖表類型：如一個測量值以次數分配表呈現、兩個測量值則以散布圖表示、三個以上的測量值以氣泡圖表現等。全視覺化地理空間視圖，對地理空間資料一目了然，協助使用者快速透析空間資料；「What does it

mean」對話視窗自動找出並解釋參數間關係；可製作具光影效果之 3D 圖表；自動繪圖功能會呈現最適合資料類型的圖表，省去反覆嘗試的步驟；內建各種精美圖表如盒形圖、熱圖、氣泡圖；工具面板中的各種圖示可讓使用者變更查詢或自由篩選、分類資料；敘述性統計選單：如最大值、最小值、平均數等可輕鬆掌握特定測量值，使用者可自行依其需求自選功能服務、即時性的視覺化資料探索和瀏覽功能，能夠快速判讀資料的重要意義。

- **使用者自主分析：**SAS 巨量資料視覺分析平台，整合了先進分析技術和各種實用功能，如快速預測功能、自動繪圖功能、「詞彙說明」話框提示、滑鼠拖曳等功能。不論是否具備統計知識，是否為統計分析人員或者具有分析技術背景，都能透過功能選單，藉此理解複雜的資料，執行強大的資料分析。可篩選各層資料進行多維度交叉分析；In-Memory SAS LASR 分析伺服器運算技術可分析巨量資料，解讀大量變數間之關聯度，任意彙總或探鑽資料及執行切割運算。

◆ 行動智慧

支援 iPad 與 Android 原生手勢功能（如拉近/拉遠、輕掃功能）貼近使用者使用習慣；其彈性排版功能為不同需求者量身打造報表內容；依連結指標圖出定義建立報表、或依選取指標符號做為篩選標準；支援安全的上線/離線模式瀏覽報表；報表可添加註解協助瞭解內容也可依電子郵件信箱附加註解的報表傳送其他相關人員；報表變更時能推送至行動裝置。

◆ SAS 高效能運算分析

SAS 高效能分析產品（SAS BASE、SAS/ETS、SAS/STAT、SAS Enterprise Miner）搭配 SAS 合作夥伴（EMC Greenplum 或 Teradata）的 Database appliance 解決方案，利用 In-Memory 處理資源，將資料庫系統視為分享記憶體之網格運算，可大量平行運算 SAS 程式碼及分析過程，快速移動大量 In-Memory 資料，以即時產出分析結果。

SAS 高效能運算分析，不但能執行高效能資料探、高效能變數簡化、並可做高效能線性迴歸分析、高效能羅吉斯迴歸分析、高效能非線性迴歸分析、甚至高效能混合線性迴歸模型、高效能資料採礦、高效能神經網路以及隨機森林決策樹、高效能計數迴歸分析、高效能嚴重程度模型。其原理是透過一連串的平行程序提供大規模資料探索，藉由此探所能力，可由巨量資料快速得到敘述統計：如平均值、最大值、最小植、全距、離散程度、集中程度等，以利集合、彙總及各層級變數一併產出。以其高效能變數縮減特性，找出解釋最大量資料變異的變數，並選擇非監督式變數。此外，HPREDUCEY 程序，可選出影響整體資料變異甚鉅的原始變術子集，執行變異數分析並簡化變數維度。

在處理巨量資料與複雜運算時系統會擷取資料並重新配置於精密的資料庫記憶體的手法，使得資料不受結構限制，快速置入記憶體，快速處理新情境案例或新運算，可幫助運用大量資料及複雜分析，解決困難的問題或過去認為無法做的窘境。

有了此種高效能運算分析技術，無需使用子資料集，可一次使用全部的資料應用在多情境案例下進行資料準備、探索與建立模型，並即時回應，以利更快速做出高精準度的分析見解或關鑑決策。

◆ SAS Mobile BI

平板電腦與智慧手機的廣泛使用，使得組織需要採用和支援行動裝置，IT 人員必需為商務使用者和管理階層提供最新的相關資訊，隨時將資訊送到使用者的行動裝置中。SAS Mobile 是 SAS Visual Analytics 解決方案的重要元件，利用 Visual Analytics 易於使用的報表設計工具，可快速的建立管理層級的報表、視覺化資訊，透過 Web 發布（Adobe PDF 檔或發布至 iPad 或 Android 行動裝置，大幅加快中要資訊的發送，促進知識分享和決策，並提供安全（VPN、SSL、HTTPS、黑名單與遠端擦除）的互動資訊，讓行動決策者能夠快速取得深入分析後的資訊。

- 報表物件的豐富互動功能，提供報表設計與資料使用的無限可能，使用者只需利用行動裝置中內建的互動功能（如縮放、滑動、內容分組、單擊瀏覽）可瀏覽、鑽探、篩選報表物件等進行互動。
- SAS Mobile BI 具備離線功能，即使沒有網路連線，也可以隨時隨地存取：此外可顯示 SAS Stored Process 的輸出，不需存取 SAS 應用程式：使用現有的資料，依據最新的資訊，為使用者提供即時分析。

肆、心得與建議

一、資料科學與資料科學家

資料本身告訴我們「發生了什麼」(WHAT)，而我們必需能夠由資料中發現「為什麼」(WHY)。以往，如何去統計和整合不同來源的資料一直是我們最大的挑戰，有太多的資訊我們無法掌握，使得我們不得不倚賴過去的經驗來做判斷。如今，使用全部的資料做分析已不成問題，許多原來需要投入的軟硬體成本已經大大降低，技術發展上也已經成熟不少，甚至有許多非常經典的實例已經出現。例如：**Amazon** 的書籍推薦系統、**Wallmart** 懂得在颶風前促銷小甜點。IT 應用系統既然已經可以負荷得了越來越多的資料，接下來要思考的是，在分析作業上是否更趨於主動，提供更深入的分析結果，不只是告訴人們 **What**，而且更有機會回答 **How** 和 **Why** 等問題。

資料科學是指在資料庫中利用各種分析方法與技術，將過去所累積的大量繁雜歷史資料，結合即時的外部或非結構資料，進行分析、歸納與整合等工作，萃取出有用的資訊，找出有意義且使用者有興趣的樣式 (**pattern**)，提供企業管理階層在進行決策時的參考依據。資料科學應用不只是清理資料、建立分析模型，應包含資料清理前的收集資料及處理資料流程，清理資料後之探索式分析、機器學習、建立資料產品。

資料科學家 (**Data scientist**) 是集合業務分析、產業顧問、溝通協調及資料科學專業於一身的人，並懂得從當今巨量排山倒海而來的非結構性資訊當中，抽絲剝繭，找到重要問題的答案。據全球最具影響力的商業雜誌「財富」報導，資料科學家是未來十年最具前瞻性的工作；全球專業經理人及工商管理者都要閱讀的「哈佛商業評論」雜誌甚至在 2012 年 10 月號指出：資料科學家是企業最誘人的職缺。因此，資料科學家絕對是未來

職場中最炙手可熱的明星職位。自 2012 年起，巨量資料的應用，已然成為全球科技趨勢的主角，而擁有高效能的商業分析平台及優質的「資料科學家」人才，幫助企業快速作出正確決策，成為企業決戰商場的兩大利器。2013 年 IDC 台灣巨量資料分析市場調查報告中顯示，台灣企業對巨量資料分析的需求在未來 3 年內將有更顯著的成長。研究機構 Gartner 預測，2015 年全球將有 440 萬個海量資料相關的 IT 工作職缺，唯目前尚未有真正以巨量資料為背景的學科，因此人才缺口恐達 3 分之 2。

「分析能力是一種結合科技、數據及經驗的整合有機體」，資料科學家的養成需整合多方面的基礎能力，包含商業觀念、統計、資訊管理、邏輯、分析等。而為了結合技術、資料，以及應用分析與服務等三方面的功能，增強巨量資料分析之跨領域整合人才培訓以及相關國際學術交流，並透過結合跨領域研究人力資源，提升研究能量。成功的巨量資料分析人員，除須對統計方法有廣泛的了解外，還須要：1. 在某些科學領域有深厚的專業知識，因而能將要解決的問題，分解成一組明確且可測試的假設。2. 具備足夠的電腦知識與程式能力，以利於使用最有效率的電腦處理技術，來管理與分析比先前更大的資料量。3. 能夠以非技術性的語言，來解釋分析結果，以達成有效率的溝通。故如何整合統計分析、資訊科學、資料視覺化的跨領域專業，建立「巨量資料分析」人才是刻不容緩的事情。

很多公司其實本身都有一些人員在分析資料，這些資料分析員表面上擔任分析的角色，但實際上只是在製作報表，他們是資料科學家嗎？或者可能進階成為資料科學家嗎？科學家這樣的工作，本來就有探索、研究的特質，若一直期待別人指示去作分析、發掘問題，沒辦法產生新的價值。但具有這種能力或專業的人才太難找尋，因為資料科學面對的範疇複雜，不是人人都有辦法同時精通資料採礦、機器學習等技術，單靠一個人要具備這麼多不同能力，也不容易，因此需透過團隊合作的方式（該業務領域

的專家、資料科學工程師 (Data Science Engineer)、統計學家或分析師)：讓每個人在一個領域很專精，其他人也涉獵一些這個領域的相關知識，彼此互補。

早在 2012 年台灣微軟瞄準此需求，和中華資料採礦協會合作，提供微軟的分析工具，並結合學界專家師資，成立「微軟資料科學研究院」，課程內容包括資料科學基礎、企業資訊整合、資料探索、數據挖掘應用等等，透過產學合作，落實企業社會責任，希望及早在台培育更多資料科學人才，提升台灣在全球巨量資料分析環境的競爭力。

2012 年起，SAS 臺灣分公司的「校園資料科學家育才計畫」，藉由 SAS 商業分析領域的深厚經驗與專業的分析技術，與全台逾 50 多間頂尖大學合作，啟動資料科學家培訓課程計畫，透過產學合作，孕育資料分析人才、提升台灣在全球巨量資料分析環境的競爭力。並且舉辦「SAS 校園資料採礦競賽－高手高手高高手」，透過提供企業去識別化資料庫實例方式，各參賽團隊在學校師長指導下，應用 SAS 的分析軟體，由學生團隊提出創意的資料採礦運用提案，參加競賽。透過此一全國性比賽，集合各校優秀學子交流切磋，互相激盪，增加資料庫採礦運用的實力，這也是開放資料 (open data) 倡導讓大眾將資料加值的作法之一。

二、大資料需要大分析 (Big Data need Big Analytics)

大資料與資料處理、資料分析有關，可以說是一個偏技術性的議題，因 MapReduce, Hadoop 等分散式巨量資料處理的問世，讓人們在資料處理上躍進了許多，但是有了這些新的處理技術，並不會讓空有巨量資料的擁有者突然變得比較聰明或能有嶄新的見解，商業智慧需經下列八種演進階段，才能將資料轉化為資訊、由挖掘知識到洞見、決策。

(1) 標準報表 — What happened? When did it happened?

- (2) 隨選報表 — How often? How many? Where?
- (3) OLAP 查詢 — Where exactly? How to find answer?
- (4) 警示 — When should I act? What actions?
- (5) 統計式分析 — Why is it happened? What opportunities am I missing?
- (6) 預測 — What if? How much? When?
- (7) 預測建模 — What will next happen?
- (8) 最佳化 — Best decision?

若沒有明確的目標與應用，巨量資料不會有任何價值，長期以來，總處負責各項數據的蒐集、分析、彙編和公布，由於是官方的統計發布窗口，因此這些資料必需全面、可靠，同時具有國際性參考價值。但統計報表僅意味提供事實，累積了多年的歷史資料，看得見的問題容易被正視，對於看不見的問題，就需要透過資料分析來突顯。如何由這些大量資料中找到有意義的資訊，快速地洞悉完整的決策資訊，勢必得尋求不同已往的資料處理解決方案。例如：Google 工程師發表的以資料分析技術預測流感趨勢的研究（Flu Trends），僅靠分析人們在網路世界中搜尋引擎上的用字而不靠基層醫師回報的流感案例，的確令學者專家跌破眼鏡，這種另類的分析用法可能不是僅靠運氣即可做得到，必需像採挖石油般探勘多次才能找到有產值的油井，若枯竭時，也需要判斷是否繼續採挖或放棄而尋找另一處油井。

在這巨量資料的時代裡，不但許多技術在改變，連觀念與做法也必需時時調整—如何達標所涉及的技术問題，理應由資訊部門來負責相關的儲存設備與網路等後端系統，甚至基礎設備；如何活用資料以助於改善決策

勢必也不能不正視；人腦固然厲害，卻獨缺電腦的擴充力。善用工具以強化能力，製作應用程式，讓人們看了能直覺瞭解的可視化作業，藉由可視化直覺地掌握趨勢，由動態進展到即時甚而到互動，適時的引進可視化分析（**Visual Analytics**）以便看到以往忽略的狀況。互動交談式分析，則允許各種模型的試誤。資料採礦分析、預測分析是以更客觀的理論導出模型。在常日的資料彙整外，可否根據歷史資料找出規則，設計出一系列的統計模型。有新的資料時，套用到模型裡，接著模型就會顯示出一個可接受的範圍，如果超過範圍，可計算差值判斷是否異常，以檢測工作中潛藏的瑕疵或失誤，降低有誤資料導致錯誤的結論，由一般的資料分析進展到趨勢分析似乎是可以考慮的手法。

三、 將資料當做策略性、核心資產來管理

總處不論是公務統計或調查統計業務，累積了多年的各種資料，可否將資料當做重要資產來管理，藉由資料整合，改善整個組織的精確資訊流程；透過資料生命週期管理，確保資料品質良好；建立單一、精確且統一的資料檢視機制，將資料整合、資料品質與管理合併為統一的開發與遞送環境，組織可充分運用資料管理程序的每一個階段，將一致、精確且即時的資料提供給組織，讓分析人員不用再費力連結散落各伺服器、各儲存體的各種結構的資料，專心致力於其統計分析領域，建立及分享深度資訊。

四、 地理視覺化報表

資料無處不在：所有數據的產製，或多或少都與「地理位置」有密切的關係，如商圈分析、銷售據點分析、**Location-Based Service App**。地圖報表是商業趨勢與地理資訊的完美結合，也就是地理資料視覺化—藉由資料驅動圖形的行為與屬性。善用圖形（**Shape**）的行為進行資料視覺化，在圖形中建立行為模式，讓圖形資料有所異動時改變圖形的外觀、面貌與

行為。地理視覺化報表，透過地圖幫助人們更容易理解空間資訊，則總處多年來以文字處理模式企圖解譯全國門牌位置空間資訊的無力感，將能有所改善。

伍、結語

國際專家 2014 年預測 IT 趨勢的五大面向：1.行動化趨勢後的企業重心調整 2.大數據成長與應用層面 3.傳統 IT 基礎建設 (Infrastructure) 的重新定義 4 企業流程優化與管理 5.資訊安全新型態威脅與防禦。 巨量資料是下世代最重要的五大趨勢之一，「資料科學家」更是因此趨勢而產生的新機會。不記得在哪個場合聽過頗精闢的見解：由知其然 (資訊報表) 到知其所以然 (行為)，由結構性資料 (資訊) 到凡走過必留下痕跡 (訊息)，由事後聰明 (資訊統計) 到事前預測 (訊息預測)，由市場區隔 (企業資訊) 到建構網路 (消費者訊息)，由數據分析 (資訊) 到視覺化 (圖像訊息)，由多樣性 (分散資料) 到整合性 (大數據)，以上這幾種趨勢點出資訊即行為，瞭解體驗到知的力量，大分析的時代就是由知獲利。

借用謝邦昌教授的名言：未來的世界除了「軟體即服務」外，更多了「資訊/智能即服務 (Information & Intelligent as a service)」。

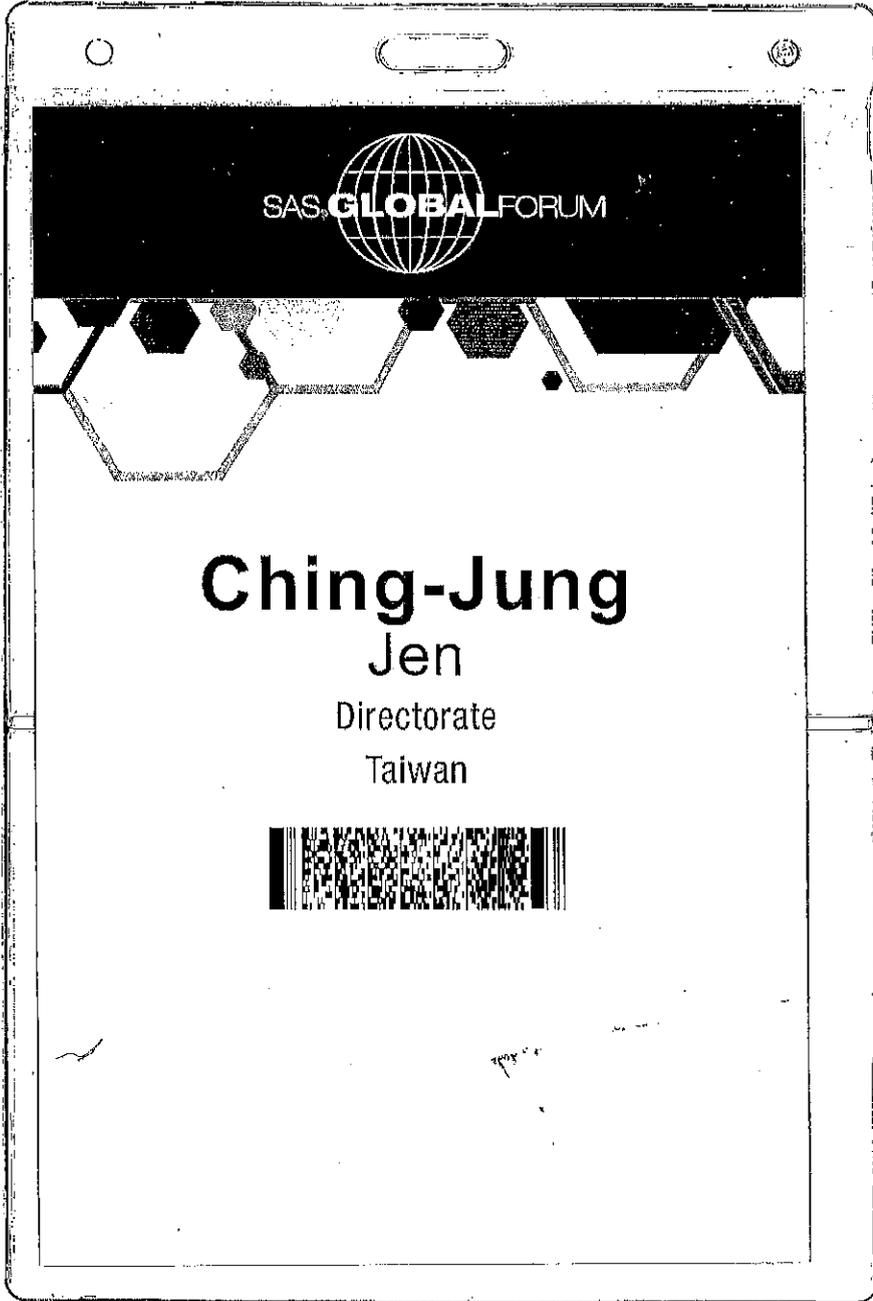
但 SAS 公司其商業標語更精準的指出：優越的分析技術，幫助人們快速獲得深度資訊；資料視覺化，看到以往忽略的現象；行動商業智慧，在移動中即時獲取資訊的能力。

巨量資料的時代已然來臨，如何不在資料的洪流中沒頂，挖掘出有價值的金礦是頗值得深思的議題。

陸、參考資料

1. <http://www.sas.com/offices/asiapacific/taiwan/>臺灣賽仕電腦。
2. <http://www.ithome.com.tw/tech/>
3. http://www.microsoft.com/taiwan/newsletter/library/edm2013/1003_SQLeDM/微軟資料科學研究院
4. 維基百科

柒、附錄



Monday Morning Sessions - March 24

NATIONAL HARBOR 2	NATIONAL HARBOR 3	NATIONAL HARBOR 4	NATIONAL HARBOR 5	NATIONAL HARBOR 6-7	NATIONAL HARBOR 10-11	NATIONAL HARBOR 12-13	CHESSAPEAKE A-C	CHESSAPEAKE D-F	CHESSAPEAKE G	CHESSAPEAKE H-I	CHESSAPEAKE J-L	CHESSAPEAKE 2-3	CHESSAPEAKE 4-6	CHESSAPEAKE 9	POTOMAC 1-3
-------------------	-------------------	-------------------	-------------------	---------------------	-----------------------	-----------------------	-----------------	-----------------	---------------	-----------------	-----------------	-----------------	-----------------	---------------	-------------

Breakout Session
 Hands-On Workshop
 Q&A/Tips
 Panel
 Live Stream

Technology Connection and Keynote Presentation: Nelson Dellis
 Convention Center - Level 2, Potomac Ballroom
 8:00 - 10:00 a.m.

4911: Modernizing our Data Strategy Lecture [B] 10:30-11:20 a.m.	1800: Seven Steps to a SAS Enterprise Value [B] 10:30-11:20 a.m.	1899: Macro Design Fund [B] 10:30-11:20 a.m.	1265: SAS Enterprise Guide Speed [B] 10:30-11:20 a.m.	1410: Something for Nothing! [B] 10:30-11:20 a.m.	1483: The Art of Clear Data Task [B] 10:30-11:20 a.m.	SAS1393: SAS Office Analytics Review [B] 10:30-11:20 a.m.	SAS1979: Case-Level Risk Analysis [B] 10:30-11:20 a.m.	1204: SAS Functions Copy [B] 10:30-11:20 a.m.	SAS1716: Leveraging SAS Visualization [B] 10:30-11:20 a.m.	SAS208: SAS Visual Analytics Report [B] 10:30-11:20 a.m.	SAS264: Reading and Writing ZIP [B] 10:30-11:20 a.m.	SAS105: So Much Software [B] 10:30-11:20 a.m.	SAS247: Smart Meter Analytical Calls [B] 10:30-11:20 a.m.	SAS341: Building Stronger Communities [B] 10:30-11:20 a.m.	SAS1467: Missing Data [B] 10:30-11:20 a.m.
345934: What's New in SAS Research [B] 11:30 a.m.-12:20 p.m.	1318: Secure SAS OIAP Codes [B] 11:30-11:50 a.m.	1271: DATA Step Merging Techniques [B] 11:30 a.m.-12:20 p.m.	SAS3385: High Memory Switch to SAS [B] 11:30 a.m.-12:20 p.m.	1444: Communicate Non-Enterprise Data [B] 11:30 a.m.-12:20 p.m.	1475: The Art of Clear Data Task [B] 11:30 a.m.-12:20 p.m.	SAS1987: SAS Office Analytics Review [B] 11:30 a.m.-12:20 p.m.	2044: Enhancing Optical Readability [B] 11:30 a.m.-12:20 p.m.	1459: Omit, Hide, My Favorite SAS Tricks [B] 11:30 a.m.-12:20 p.m.	SAS156: Putting on the Biz New Ways [B] 11:30 a.m.-12:20 p.m.	SAS120: Getting the Most Out of SAS [B] 11:30 a.m.-12:20 p.m.	SAS264: Reading and Writing ZIP [B] 11:30 a.m.-12:20 p.m.	SAS107: The Latest in SAS Enterprise [B] 11:30 a.m.-12:20 p.m.	SAS371: What's New in SAS Enterprise [B] 11:30 a.m.-12:20 p.m.	2068: What Benefits Can a Hospital Achieve [B] 11:30 a.m.-12:20 p.m.	SAS102: Consumer Research Tools [B] 11:30 a.m.-12:20 p.m.
1835: Real-Time Market Monitoring Dashboard [B] 11:30 a.m.-12:10 p.m.	1835: Real-Time Market Monitoring Dashboard [B] 11:30 a.m.-12:10 p.m.	1835: Real-Time Market Monitoring Dashboard [B] 11:30 a.m.-12:10 p.m.	1835: Real-Time Market Monitoring Dashboard [B] 11:30 a.m.-12:10 p.m.	1835: Real-Time Market Monitoring Dashboard [B] 11:30 a.m.-12:10 p.m.	1835: Real-Time Market Monitoring Dashboard [B] 11:30 a.m.-12:10 p.m.	1835: Real-Time Market Monitoring Dashboard [B] 11:30 a.m.-12:10 p.m.	1835: Real-Time Market Monitoring Dashboard [B] 11:30 a.m.-12:10 p.m.	1835: Real-Time Market Monitoring Dashboard [B] 11:30 a.m.-12:10 p.m.	1835: Real-Time Market Monitoring Dashboard [B] 11:30 a.m.-12:10 p.m.	1835: Real-Time Market Monitoring Dashboard [B] 11:30 a.m.-12:10 p.m.	1835: Real-Time Market Monitoring Dashboard [B] 11:30 a.m.-12:10 p.m.	1835: Real-Time Market Monitoring Dashboard [B] 11:30 a.m.-12:10 p.m.	1835: Real-Time Market Monitoring Dashboard [B] 11:30 a.m.-12:10 p.m.	1835: Real-Time Market Monitoring Dashboard [B] 11:30 a.m.-12:10 p.m.	1835: Real-Time Market Monitoring Dashboard [B] 11:30 a.m.-12:10 p.m.

Networking Opportunities

Lunch, 12:30 - 1:30 p.m.
 Convention Center - Level 1, Prince George's Exhibition Hall B, C, Level 2, Chesapeake Lobby
 Use your voucher at kiosks located throughout the conference to self-select your meal.
 The voucher will be provided at registration.

A. 80% save in
 80% change
 A. schedule

CTL + 2

1. Chris Howard's Blog
 2. Communities - SAS on the
 3. Low Tension Search
 4. SAS Download → EG → Support

SAS SUPPORT AND DEMO AREA: Monday 10:00 a.m. - 4:50 p.m., 6:00 - 7:30 p.m.
 CODE DOCTORS: Monday 10:30 a.m. - 12:30 p.m., 2:00 - 4:00 p.m.
 e: SAS Meet

Complete the session survey for the sessions you attend at sasglobalforum.com/session-survey.

TELL US WHAT YOU THINK



Monday Afternoon Sessions - March 24

Breakout Session
 Hands-On Workshop
 Quick Tip
 Panel
 Feature
 Live Stream

NATIONAL HARBOR 2	NATIONAL HARBOR 3	NATIONAL HARBOR 4	NATIONAL HARBOR 5	NATIONAL HARBOR 6-7	NATIONAL HARBOR 10-11	NATIONAL HARBOR 12-13	CHESSAPEAKE A-C	CHESSAPEAKE D-F	CHESSAPEAKE G	CHESSAPEAKE H1	CHESSAPEAKE J-L	CHESSAPEAKE 2-3	CHESSAPEAKE 4-6	CHESSAPEAKE 9	POTOMAC 1-3
5AS256: Excel Tips Sharon U 1:30-1:50 pm	1634: SAS Visual Analytics Publishers Self U 1:30-1:50 pm	1401: Reading in Data Priority Doreen U 1:30-1:50 pm	1468: Parallel Guide to Publishing Doreen U 1:30-1:50 pm	5AS168: Extending Key Concepts Marty U 1:30-2:00 pm	2126: SAS Enterprise Guide 5.1 Felix U 1:30-2:00 pm	5AS1523: Data Mining Luan Speed U 1:30-3:10 pm	1656: Multivariate Time Series Modeling Felix U 1:30-2:00 pm	5AS1984: Diagnostic and Inference Tools Gabe U 1:30-1:50 pm	5AS1728: Living in Perfect Harmony Suzanne U 1:30-2:00 pm	1545: Association Mining of Big Data Meredith, Kellen, Steve U 1:30-1:50 pm	5AS142: Security Scenario Schaefer U 1:30-1:50 pm	1694: Grid-White They Don't Tell You Meredith, Wenzel U 2:30-2:40 pm	1456: ACR/CT/HARCI Reading Lester U 1:30-2:00 pm	5AS1523: Work Area Organization at Major Heon U 1:20-2:20 pm	5AS2205: Getting Started with SAS 1:30-2:20 pm
5AS247: Big Data Experiential Mabel U 2:30-2:50 pm	1486: How to Be a Data Scientist A. Data Scientist U 1:30-2:20 pm	1316: Getting the Most Out of SAS Dana U 2:00-2:50 pm	1326: Pre-Workshop Flip-Flip Your Data Helen U 2:30-3:20 pm	5AS1282: Building Your Own SAS Cloud Villers U 2:30-2:50 pm	1627: Grasping the Made Easy With ODS Drew U 3:30-4:20 pm	5AS1421: SAS Visual Analytics Resant U 3:30-5:10 pm	1775: Application of Text Mining in Wealth Dale U 2:00-2:30 pm	1731: Ordering Columns in a SAS Clyde U 2:00-2:30 pm	5AS153: SAS Format Optimization Wade U 2:00-2:30 pm	1581: Building Genes Suzanne U 2:00-2:30 pm	1811: NIBP Pharmacy Surveillance System Meredith U 2:00-2:30 pm	1721: User-Friendly SAS Grid U 2:30-3:00 pm	1784: How to Separate Regular Press Lester U 2:00-2:20 pm	5AS1523: SAS Visual Analytics Resant U 3:30-5:10 pm	5AS1523: SAS Visual Analytics Resant U 3:30-5:10 pm
5AS238: Selling Over the ACROSS Hurdle Zander U 3:00-3:00 pm	1324: Uncover the Most Common Analytics Hurdle Amanda U 3:00-3:20 pm	1602: Develop Highly Interactive Web Inbar U 3:30-3:40 pm	1732: A Note on Type Conversions Chapin U 4:10-4:20 pm	5AS101: DASH/Team SAS Enterprise Guide M. Lin U 3:30-4:00 pm	1730: RROC MBOUATI M. Lin U 3:30-4:00 pm	1618: Effectively Utilizing Laptops U 4:30-5:00 pm	5AS181: SAS/STAT 13.1 Round-Up Bedinger U 2:30-3:30 pm	1761: Making it Happen Suzanne U 2:00-2:30 pm	2085: Leveraging Mathematika Wendell U 2:00-2:30 pm	1713: The Role of Customer Response Yves U 2:30-2:40 pm	2082: How to Leverage SAS Visual Analytics Comet, Lester, Stephanie U 2:30-3:00 pm	1815: A SAS Style Suzanne U 3:00-3:30 pm	5AS274: Selection Models to Retelling Wade U 2:30-2:40 pm	1636: Application of Survey Sampling De U 3:00-3:20 pm	1487: Using SAS ODS Graphics Kendall U 4:30-5:20 pm
1428: Before You Get Shared Catherine U 4:30-5:00 pm	2273: SAS and Java Application Dunaway, Wang U 3:00-4:00 pm	1645: Speed Dating Loping U 4:20-4:30 pm	1695: Best Practices in SAS Enterprise Guide U 4:30-4:50 pm	1838: Creating Formats on the Fly Doreen U 3:50-4:10 pm	1839: Creating Formats on the Fly Doreen U 3:50-4:10 pm	5AS181: SAS/STAT 13.1 Round-Up Bedinger U 2:30-3:30 pm	1820: Internal Credit Ratings Gabe U 3:00-3:30 pm	2030: Developing the Code to Evaluate Applicant Creditworthiness U 4:00-4:20 pm	5AS276: SAS Decision Management Suzanne U 2:40-3:30 pm	1837: The Use of Analytics for Insurance Kevin U 3:00-3:50 pm	5AS346: Create Custom Graphics Doreen U 3:00-3:50 pm	1492: Introduction to Fuzzy Models Amanda U 2:40-3:00 pm	1636: Application of Survey Sampling De U 3:00-3:20 pm	1622: The SAS/STAT Algorithm Raza U 3:20-3:40 pm	1874: A Mental Health and Risk Behavior Suzanne, Gabe U 3:00-4:00 pm
5AS232: The Desert and the Dunes Amanda U 3:00-3:50 pm	1726: Developing Web Applications Wade U 4:30-4:50 pm	1440: What's Hot Missing U 4:30-4:50 pm	1575: The SAS Enterprise Guide U 4:30-4:50 pm	1495: Jazz It Up a Little With Formats Doreen U 4:30-5:00 pm	1495: Jazz It Up a Little With Formats Doreen U 4:30-5:00 pm	1691: SAS/STAT 13.1 Round-Up Bedinger U 2:30-3:30 pm	1691: SAS/STAT 13.1 Round-Up Bedinger U 2:30-3:30 pm	1691: SAS/STAT 13.1 Round-Up Bedinger U 2:30-3:30 pm	1691: SAS/STAT 13.1 Round-Up Bedinger U 2:30-3:30 pm	1691: SAS/STAT 13.1 Round-Up Bedinger U 2:30-3:30 pm	1691: SAS/STAT 13.1 Round-Up Bedinger U 2:30-3:30 pm	1691: SAS/STAT 13.1 Round-Up Bedinger U 2:30-3:30 pm	1691: SAS/STAT 13.1 Round-Up Bedinger U 2:30-3:30 pm	1691: SAS/STAT 13.1 Round-Up Bedinger U 2:30-3:30 pm	1691: SAS/STAT 13.1 Round-Up Bedinger U 2:30-3:30 pm
5AS233: The Desert and the Dunes Amanda U 3:00-3:50 pm	1726: Developing Web Applications Wade U 4:30-4:50 pm	1440: What's Hot Missing U 4:30-4:50 pm	1575: The SAS Enterprise Guide U 4:30-4:50 pm	1495: Jazz It Up a Little With Formats Doreen U 4:30-5:00 pm	1495: Jazz It Up a Little With Formats Doreen U 4:30-5:00 pm	1691: SAS/STAT 13.1 Round-Up Bedinger U 2:30-3:30 pm	1691: SAS/STAT 13.1 Round-Up Bedinger U 2:30-3:30 pm	1691: SAS/STAT 13.1 Round-Up Bedinger U 2:30-3:30 pm	1691: SAS/STAT 13.1 Round-Up Bedinger U 2:30-3:30 pm	1691: SAS/STAT 13.1 Round-Up Bedinger U 2:30-3:30 pm	1691: SAS/STAT 13.1 Round-Up Bedinger U 2:30-3:30 pm	1691: SAS/STAT 13.1 Round-Up Bedinger U 2:30-3:30 pm	1691: SAS/STAT 13.1 Round-Up Bedinger U 2:30-3:30 pm	1691: SAS/STAT 13.1 Round-Up Bedinger U 2:30-3:30 pm	1691: SAS/STAT 13.1 Round-Up Bedinger U 2:30-3:30 pm

@ Jennifer
 @ Sarah, Deborah
 @ reuse
 Network Loop Consulting
 Joshua M. Konstantin

Miker Hosted by SAS Customer Loyalty, 6:00 - 7:30 pm.
 Convention Center - Level 1, Prince Georges Exhibition Hall D/E
 SAS servers as your host for this Monday evening mixer in the SAS Support and Demo Area
 This is where we celebrate SAS users! Mix and mingle with other conference attendees
 and enjoy some great food, drinks and fun entertainment.

SAS SUPPORT AND DEMO AREA: Monday 10:30 a.m. - 4:00 p.m., 6:00 - 7:30 p.m.
 CODE DOCTORS: Monday 10:30 a.m. - 12:30 p.m., 2:00 - 4:00 p.m.

Schedule subject to change. Please visit sasglobalforum.com or download the conference app from your device's app store for the most up-to-date schedule.

Complete the session survey for the sessions you attend at sasglobalforum.com/session-survey



TELL US WHAT YOU THINK

Tuesday Afternoon Sessions - March 25

NATIONAL HARBOR 2	NATIONAL HARBOR 3	NATIONAL HARBOR 4	NATIONAL HARBOR 5	NATIONAL HARBOR 6-7	NATIONAL HARBOR 10-11	NATIONAL HARBOR 12-13	CHEESAPEAKE A-C	CHEESAPEAKE D-F	CHEESAPEAKE G	CHEESAPEAKE H-I	CHEESAPEAKE J-L	CHEESAPEAKE 2-3	CHEESAPEAKE 4-6	CHEESAPEAKE 9	POTOMAC 1-3
-------------------	-------------------	-------------------	-------------------	---------------------	-----------------------	-----------------------	-----------------	-----------------	---------------	-----------------	-----------------	-----------------	-----------------	---------------	-------------

Keynote Session: Jason Dorsey
 Convention Center - Level 2, Potomac Ballroom
 12:45 - 2:00 p.m.

AS3105: Best Practices for Implementing SAS Code 09:25-10:25 p.m. Green [B]	AS3115: SAS 9.4 Web 17:15: Counting Days of the Week 09:30-10:30 p.m. Green [B]	AS3115: SAS 9.4 Web 17:15: Counting Days of the Week 09:30-10:30 p.m. Green [B]	AS3115: SAS 9.4 Web 17:15: Counting Days of the Week 09:30-10:30 p.m. Green [B]	AS3115: SAS 9.4 Web 17:15: Counting Days of the Week 09:30-10:30 p.m. Green [B]	AS3115: SAS 9.4 Web 17:15: Counting Days of the Week 09:30-10:30 p.m. Green [B]	AS3115: SAS 9.4 Web 17:15: Counting Days of the Week 09:30-10:30 p.m. Green [B]	AS3115: SAS 9.4 Web 17:15: Counting Days of the Week 09:30-10:30 p.m. Green [B]	AS3115: SAS 9.4 Web 17:15: Counting Days of the Week 09:30-10:30 p.m. Green [B]	AS3115: SAS 9.4 Web 17:15: Counting Days of the Week 09:30-10:30 p.m. Green [B]	AS3115: SAS 9.4 Web 17:15: Counting Days of the Week 09:30-10:30 p.m. Green [B]	AS3115: SAS 9.4 Web 17:15: Counting Days of the Week 09:30-10:30 p.m. Green [B]	AS3115: SAS 9.4 Web 17:15: Counting Days of the Week 09:30-10:30 p.m. Green [B]	AS3115: SAS 9.4 Web 17:15: Counting Days of the Week 09:30-10:30 p.m. Green [B]	AS3115: SAS 9.4 Web 17:15: Counting Days of the Week 09:30-10:30 p.m. Green [B]	AS3115: SAS 9.4 Web 17:15: Counting Days of the Week 09:30-10:30 p.m. Green [B]
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---

Networking Opportunities

- Charity Event Red Wagon Brigade, 5:00 - 6:30 p.m.
Convention Center - Level 1, Prince George's Exhibition Hall A
- Roll up your sleeves and build wagons, decorate them and fill them with STEM (Science, Technology, Engineering and Math) related books and toys! This is a great chance to do something special with your SAS comrades while having fun and giving back to the DC community.
- Kick Back Party, 9:00 p.m. - Midnight
Convention Center - Level 2, Maryland Ballroom
- Plan on kicking back while you enjoy drinks and a night of fun with your conference friends.

SAS SUPPORT AND DEMO AREA: Tuesday 10:30 a.m. - 5:00 p.m.
 CODE DOCTORS: Tuesday 10:30 a.m. - Noon, 2:00 - 4:00 p.m.

Schedule subject to change. Please visit sasglobal.com or download the conference app from your device's app store for the most up-to-date schedule.

Complete the session survey for the sessions you attend at sasglobal.com/session-survey.

TELL US WHAT YOU THINK



Wednesday Sessions - March 26

NATIONAL HARBOR 2	NATIONAL HARBOR 3	NATIONAL HARBOR 4	NATIONAL HARBOR 5	NATIONAL HARBOR 6-7	NATIONAL HARBOR 10-11	NATIONAL HARBOR 12-13	CHESSAPEAKE A-C	CHESSAPEAKE D-F	CHESSAPEAKE G	CHESSAPEAKE H-I	CHESSAPEAKE J-L	CHESSAPEAKE 2-3	CHESSAPEAKE 4-6	CHESSAPEAKE 9	POTOMAC 1-3
-------------------	-------------------	-------------------	-------------------	---------------------	-----------------------	-----------------------	-----------------	-----------------	---------------	-----------------	-----------------	-----------------	-----------------	---------------	-------------

General Session and Keynote Presentation: Sebastian Wernicke
 Convention Center - Level 2, Potomac Ballroom

5361: Ask Your SAS Data 19:30-20:00 am	1886: Games for Merchies 9:30-9:30 am	1240: Revamped and Hard-Shell PROC SQL User [E] 10:30-11:30 am	SAS096: Up Your Game With GenX Analytics Maple [E] 9:30-9:30 am	1341: Where in the World Are SAS/GENMAP Maps? Heller [E] 9:30-9:30 am	1482: The SAS Help Object Library (Time to find) Dierckx [E] 9:30-10:00 am	SAS1423: SAS Work- shop Data Management Richardson [E] 9:30-10:00 am	SAS193: Effective Risk Aggregation Chen [E] 9:30-9:30 am	SAS045: From Terrible to Better: Exploring Advanced Analytics 9:30-9:30 am	1277: Adding Serial Numbers to SQL Data Schrier [E] 9:30-9:30 am	1261: Using SAS to Evaluate Leifer [E] 9:30-9:30 am	1311: Time Contour Plots [E] Curtis [E] 9:30-9:30 am	2163: Retail Industry Solution: Informal Roundtable [E] 9:30-9:50 am	SAS1781: Simplifying Data Integration Farr [E] 9:30-9:50 am	1737: PROC STREAM and SAS Server Pages Henderson [E] 9:30-9:30 am	1278: Analysis of Data With Overdispersion Moral [E] 9:30-9:30 am
44: PROC MIW: Let SAS Do the Frank 9:30-10:00 am	1493: Conditional Ex- ecution "Switch Path" Shen [E] 9:30-10:00 am	1798: A Powerful SAS User's PROC EXPORT Apprentice, Intermediate [E] 11:30-11:50 am	1828: Integrated Big Data: Hadoop + DBMS + Conquering [E] 10:00-10:50 am	1342: Extreme SAS Reporting II Heller [E] 9:30-9:30 am	SAS227: Getting Started With the SAS/MIW Widder [E] 11:30 am-12:40 pm	SAS004: SAS Practice Assess Maintenance Hickel [E] 10:00-10:50 am	SAS004: SAS Practice Assess Maintenance Hickel [E] 10:00-10:50 am	1345: The end of the Road? [E] 10:00-10:20 am	SAS174: Big Digital Data, Analytic Visualization Gower [E] 10:00-10:50 am	1823: SAS Solutions to Identifying Hospital Risk [E] 10:00-10:20 am	SAS098: Unlock the Power of SAS Visual Analytics Beebe [E] 11:00-11:50 am	1824: SAS Diagnostics [E] 10:20-10:40 am	SAS174: Big Digital Data, Analytic Visualization Gower [E] 10:00-10:50 am	1823: SAS Solutions to Identifying Hospital Risk [E] 10:00-10:20 am	1686: Customer Per- ception and Health Map [E] 10:00-10:20 am
99: Combined SAS SAS Graphics 20:10-8 am	SAS021: More Than Maps: Location Chile [E] 10:00-11:50 am	SAS143: Designing for the Mobile Workforce Hickel [E] 11:00-11:20 am	1489: Custom BI Tools Using SAS Stored Calculations [E] 11:30-11:40 am	1549: Build Your Men- to with PROC CONTENTS Heller [E] 9:40-9:50 am	1673: Enhance the ODS HTML Output Hickel [E] 9:30-10:00 am	1300: Model Variable Selection Using Bootstrap Chen [E] 11:00-11:50 am	SAS046: High-Perfor- mance Forecasting Duniger [E] 10:20-10:40 am	1539: SAS Grid Man- agement [E] 10:40-11:00 am	SAS174: Big Digital Data, Analytic Visualization Gower [E] 10:00-10:50 am	SAS098: Unlock the Power of SAS Visual Analytics Beebe [E] 11:00-11:50 am	1824: SAS Diagnostics [E] 10:20-10:40 am	SAS154: Accelerated SAS to Hadoop Jucker [E] 10:30-11:20 am	SAS174: Big Digital Data, Analytic Visualization Gower [E] 10:00-10:50 am	1823: SAS Solutions to Identifying Hospital Risk [E] 10:00-10:20 am	SAS184: What's New in SAS Merchandise Murray [E] 11:00-11:30 am
48: A Framework for SAS for ETL 10:15-10:30 am	1706: Gen-Gen-Gen-Gen- Gen-Gen-Gen-Gen-Gen Lauer [E] 11:30-11:50 am	SAS046: Empowering the SAS Programmer Marsell [E] 11:30 am-1:00 pm	1489: Custom BI Tools Using SAS Stored Calculations [E] 11:30-11:40 am	1549: Build Your Men- to with PROC CONTENTS Heller [E] 9:40-9:50 am	1673: Enhance the ODS HTML Output Hickel [E] 9:30-10:00 am	1300: Model Variable Selection Using Bootstrap Chen [E] 11:00-11:50 am	SAS046: High-Perfor- mance Forecasting Duniger [E] 10:20-10:40 am	1539: SAS Grid Man- agement [E] 10:40-11:00 am	SAS174: Big Digital Data, Analytic Visualization Gower [E] 10:00-10:50 am	SAS098: Unlock the Power of SAS Visual Analytics Beebe [E] 11:00-11:50 am	1824: SAS Diagnostics [E] 10:20-10:40 am	SAS154: Accelerated SAS to Hadoop Jucker [E] 10:30-11:20 am	SAS174: Big Digital Data, Analytic Visualization Gower [E] 10:00-10:50 am	1823: SAS Solutions to Identifying Hospital Risk [E] 10:00-10:20 am	SAS184: What's New in SAS Merchandise Murray [E] 11:00-11:30 am
SAS046: Empowering the SAS Programmer Marsell [E] 11:30 am-1:00 pm	SAS046: Empowering the SAS Programmer Marsell [E] 11:30 am-1:00 pm	SAS046: Empowering the SAS Programmer Marsell [E] 11:30 am-1:00 pm	1489: Custom BI Tools Using SAS Stored Calculations [E] 11:30-11:40 am	1549: Build Your Men- to with PROC CONTENTS Heller [E] 9:40-9:50 am	1673: Enhance the ODS HTML Output Hickel [E] 9:30-10:00 am	1300: Model Variable Selection Using Bootstrap Chen [E] 11:00-11:50 am	SAS046: High-Perfor- mance Forecasting Duniger [E] 10:20-10:40 am	1539: SAS Grid Man- agement [E] 10:40-11:00 am	SAS174: Big Digital Data, Analytic Visualization Gower [E] 10:00-10:50 am	SAS098: Unlock the Power of SAS Visual Analytics Beebe [E] 11:00-11:50 am	1824: SAS Diagnostics [E] 10:20-10:40 am	SAS154: Accelerated SAS to Hadoop Jucker [E] 10:30-11:20 am	SAS174: Big Digital Data, Analytic Visualization Gower [E] 10:00-10:50 am	1823: SAS Solutions to Identifying Hospital Risk [E] 10:00-10:20 am	SAS184: What's New in SAS Merchandise Murray [E] 11:00-11:30 am
SAS046: Empowering the SAS Programmer Marsell [E] 11:30 am-1:00 pm	SAS046: Empowering the SAS Programmer Marsell [E] 11:30 am-1:00 pm	SAS046: Empowering the SAS Programmer Marsell [E] 11:30 am-1:00 pm	1489: Custom BI Tools Using SAS Stored Calculations [E] 11:30-11:40 am	1549: Build Your Men- to with PROC CONTENTS Heller [E] 9:40-9:50 am	1673: Enhance the ODS HTML Output Hickel [E] 9:30-10:00 am	1300: Model Variable Selection Using Bootstrap Chen [E] 11:00-11:50 am	SAS046: High-Perfor- mance Forecasting Duniger [E] 10:20-10:40 am	1539: SAS Grid Man- agement [E] 10:40-11:00 am	SAS174: Big Digital Data, Analytic Visualization Gower [E] 10:00-10:50 am	SAS098: Unlock the Power of SAS Visual Analytics Beebe [E] 11:00-11:50 am	1824: SAS Diagnostics [E] 10:20-10:40 am	SAS154: Accelerated SAS to Hadoop Jucker [E] 10:30-11:20 am	SAS174: Big Digital Data, Analytic Visualization Gower [E] 10:00-10:50 am	1823: SAS Solutions to Identifying Hospital Risk [E] 10:00-10:20 am	SAS184: What's New in SAS Merchandise Murray [E] 11:00-11:30 am

Wednesday Conference Tutorials - March 26

Convention Center - Level 1, National Harbor Rooms

NATIONAL HARBOR 2	NATIONAL HARBOR 3	NATIONAL HARBOR 6-7	NATIONAL HARBOR 10-11	NATIONAL HARBOR 12-13	CHESSAPEAKE A-C	CHESSAPEAKE D-F	CHESSAPEAKE G	CHESSAPEAKE H-I	CHESSAPEAKE J-L	CHESSAPEAKE 2-3	CHESSAPEAKE 4-6	CHESSAPEAKE 9	POTOMAC 1-3
-------------------	-------------------	---------------------	-----------------------	-----------------------	-----------------	-----------------	---------------	-----------------	-----------------	-----------------	-----------------	---------------	-------------

General Session and Keynote Presentation: Sebastian Wernicke
 Convention Center - Level 2, Potomac Ballroom

2239: Building Dynamic Programs Engineer 1:30-4:30 pm	2234: Introduction to Social Network Analysis Cedrick La Villier 1:30-4:30 pm	2234: Reporting Your Way - ODS LAYOUT Lund 1:30-4:30 pm	1646: Resolving Human Machine Behavior Marsell [E] 11:30-12:30 pm	1592: Mobile Reporting at the Edge of Central Finance [E] 10:00-10:30 am	1657: Using PROC GENMOD and PROC REG Zohel [E] 10:30-10:40 am	1765: Gen Reporting: Integrating ArcGIS Maps Chen [E] 10:40-11:00 am	SAS008: Better Together: Best Practices Muller [E] 11:00-11:50 am	SAS005: Financial Crimes Compliance: Track Flow [E] 11:00-12:00 pm	1300: Model Variable Selection Using Bootstrap Chen [E] 11:00-11:50 am	SAS193: Effective Risk Aggregation Chen [E] 9:30-9:30 am	1345: The end of the Road? [E] 10:00-10:20 am	SAS174: Big Digital Data, Analytic Visualization Gower [E] 10:00-10:50 am	1823: SAS Solutions to Identifying Hospital Risk [E] 10:00-10:20 am	SAS098: Unlock the Power of SAS Visual Analytics Beebe [E] 11:00-11:50 am	1824: SAS Diagnostics [E] 10:20-10:40 am
--	--	--	--	---	--	---	--	---	---	---	---	--	--	--	--

The tutorials are extra-fee events and require registration. Check with Conference Registration to see if seats are still available. Tutorials provide a more in-depth discussion than what is possible in a session.

Complete the session survey for the sessions you attend at sasglobalforum.com/session-survey.



601

[3]: 通常標示網路硬碟總容量，或具有大容量的儲存媒介之儲存容量時使用。

名字	縮寫	次方
千位元組	KB	10 ³
兆位元組	MB	10 ⁶
吉位元組	GB	10 ⁹
太位元組	TB	10 ¹²
拍位元組	PB	10 ¹⁵
艾位元組	EB	10 ¹⁸
澤位元組	ZB	10 ²¹
堯位元組	YB	10 ²⁴

[4]. 原文：Big data are high volume, high velocity, and/or high variety information assets that require new forms of processing to enable enhanced decision making, insight discovery and process optimization.

在一份 2001 年的研究與相關的演講中，麥塔集團（META Group 現為高德納）分析員道格·萊尼（Doug Laney）指出資料增長的挑戰和機遇有三個方向：量（Volume，資料大小）、速（Velocity，資料輸入輸出的速度）與多變（Variety，多樣性），合稱「3V」或「3Vs」。高德納與現在大部份大資料產業中的公司，都繼續使用 3V 來描述大資料。2012 年修改了對大資料的定義：「大資料是大量、高速、及/或多變的資訊資產，它需要新型的處理方式去促成更強的決策能力、洞察力與最佳化處理。」另外，有機構在 3V 之外定義了第 4 個 V：真實性（Veracity）為第四特點。

Global Information Storage Capacity

in optimally compressed bytes

2007 **ANALOG**

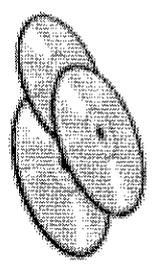
19 exabytes

- Paper, film, audiotape and vinyl: 6%
- Analog videotapes (VHS, etc): 94% **ANALOG** 
- Portable media, flash drives: 2% **DIGITAL** 
- Portable hard disks: 2.4%
- CDs and minidisks: 6.8%

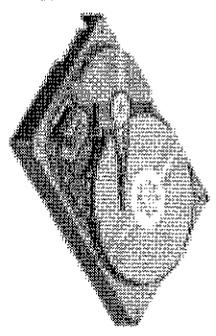
- Computer servers and mainframes: 8.9%

- Digital tape: 11.8%

- DVD/Blu-ray: 22.8%



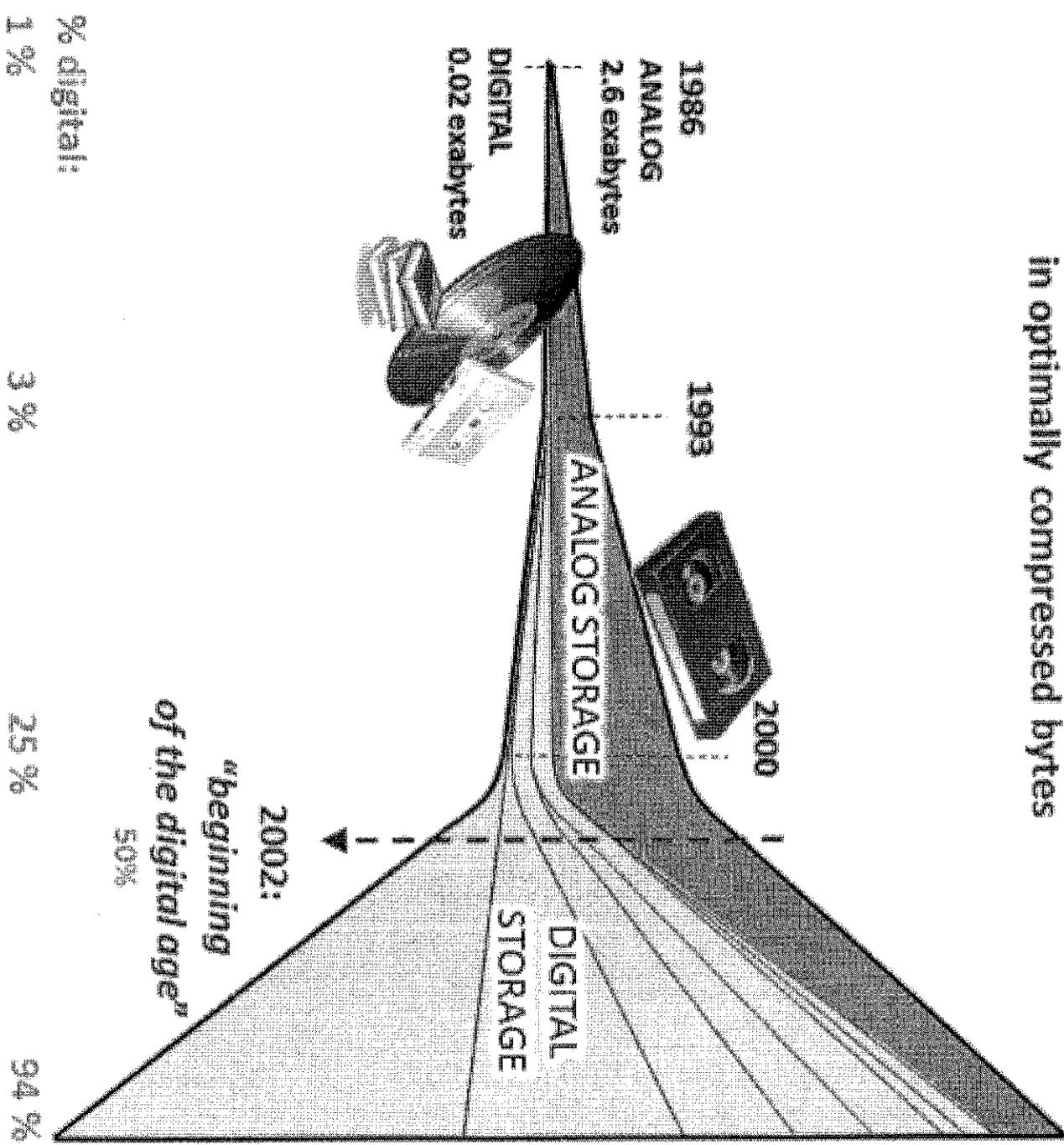
- PC hard disks: 44.5%



- Others: < 1% (incl. chip cards, memory cards, floppy disks, mobile phones, PDAs, cameras/camcorders, video games)

DIGITAL

280 exabytes



% digital:

1%

3%

25%

94%

Source: Hilbert, M., & López, P. (2011). The World's Technological Capacity to Store, Communicate, and Compute Information. *Science*, 332(6025), 60–65. <http://www.martinhilbert.net/WorldInfoCapacity.html>